Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation

Hideki Tanizawa, Osamu Iwasaki, Atsunari Tanaka, Joseph R. Capizzi, Priyankara Wickramasinghe, Mihee Lee, Zhiyan Fu and Ken-ichi Noma*

The Wistar Institute, Philadelphia, Pennsylvania, USA

ABSTRACT

We have comprehensively mapped long-range associations between chromosomal regions throughout the fission yeast genome using the latest genomics approach that combines next generation sequencing and chromosome conformation capture (3C). Our relatively simple approach, referred to as enrichment of ligation products (ELP), involves digestion of the 3C sample with a 4 bp cutter and self-ligation, achieving a resolution of 20 kb. It recaptures previously characterized genome organizations and also identifies new and important interactions. We have modeled the 3D structure of the entire fission yeast genome and have explored the functional relationships between the global genome organization and transcriptional regulation. We find significant associations among highly transcribed genes. Moreover, we demonstrate that genes co-regulated during the cell cycle tend to associate with one another when activated. Remarkably, functionally defined genes derived from particular gene ontology groups tend to associate in a statistically significant manner. Those significantly associating genes frequently contain the same DNA motifs at their promoter regions, suggesting that potential transcription factors binding to these motifs are involved in defining the associations among those genes. Our study suggests the presence of a global genome organization in fission yeast that is functionally similar to the recently proposed mammalian transcription factory.

INTRODUCTION

Eukaryotic genomes are non-randomly organized in the nucleus. It is becoming clear that intra-nuclear positions of genomic loci are influenced by various nuclear processes including transcription, replication and repair (1). It is well known that the ribosomal genes (rDNA repeats) are transcribed by RNA polymerase (Pol) I in the nucleolus. Moreover, it has been shown that Pol III genes such as tRNA genes are clustered at or near the nucleolus in yeasts, suggesting that Pol III transcription likely occurs in a subnuclear domain (2,3). It has been proposed that Pol II gene transcription involves higher-order genome organization associated with 'transcription factories' which accumulate Pol II transcription machinery for gene transcription (4–7). It has recently been suggested that transcription factors are involved in the association of genes with these transcription factories (8). However, how transcription factories function remains unclear, partly because they have been studied in complex mammalian cells. Studying the factories in a model organism with a much simpler genome can facilitate understanding of the role of transcription factories with regard to transcriptional regulation.

Fluorescent *in situ* hybridization (FISH) has been used to analyze nuclear localization of genomic loci at a global level, but a relatively new approach, chromosome conformation capture (3C), now allows us to investigate physical associations between specific genomic loci (9). The use of the 3C method has triggered development of several additional genome-wide approaches including 4C and 5C (10–12). It has recently been reported that 3C combined with next-generation DNA sequencing, referred to as Hi-C, can be used to comprehensively map genomic associations (13). Application of the Hi-C method to the human genome has identified genomic associations at a resolution of 1 Mb, and has shown that the human genome is segregated into two compartments corresponding to open and closed chromatin. We hypothesized that the latest genomics approach was likely to provide much higher-resolution if applied to a model organism carrying a small genome. Indeed, the similar method applied to budding yeast significantly

increased the resolution of mapped genomic associations (14,15).

The fission yeast *Schizosaccharomyces pombe* offers an excellent model system to investigate the organization of a functional genome. Its genome is ~14 Mb, consisting of ~5000 genes located on only three chromosomes, with an organization and composition similar to higher eukaryotes (16). For example, its genome contains large stretches of heterochromatin at centromeres and subtelomeres (17). We have previously shown that the fission yeast genome displays a specific functional architecture within the nucleus (2,18).

In this study, we utilize the latest genomic approach combining the 3C and next-generation DNA sequencing to gain insights into functional relationships between the global genome organization and transcriptional regulation in the model organism fission yeast. Our analyses have revealed significant associations between highly transcribed genes, between co-regulated genes during cell-cycle progression, and between functionally related genes derived from particular gene ontology groups. Our study identifies inter- and intra-chromosomal interactions providing further evidence for a mechanism of functional genome organization that supports gene expression in a structure similar to the transcription factory described in mammals.

MATERIALS AND METHODS

3C in fission yeast

3C analysis was performed as described previously (9) with modifications. Briefly, fission yeast cells (~7×10^8 cell) were digested by Zymolyase 100T at 30°C for 10 min, and then cross-linked with 4% paraformaldehyde at 18°C for 30 min. The fixed sample was treated with HindIII at 37°C for 2 h and then diluted 20 times with T4 DNA ligase buffer, followed by DNA ligation at 16°C for 70 min. To prepare the random ligation (RL) control sample, genomic DNA was first purified from the wild-type fission yeast strain used in 3C analysis. The genomic DNA was completely digested by HindIII at 37°C for 2 h, followed by DNA ligation. The 3C and RL samples were further subjected to the following sample preparation processes for Illumina paired end sequencing.

Enrichment of ligation products method

Eight micro grams of 3C and RL samples were digested by BfuCI at 37°C for 1 h. The resultant samples were diluted 1:10 with a T4 DNA ligase buffer and subjected to DNA ligation at 16°C for >8 h. The DNA samples were then digested by HindIII at 37°C for 2 h. The purified ELP samples were sequenced by an Illumina Genome Analyzer II. The obtained sequences have been deposited at NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra/) under the accession # SRP002804.

Physical proximity analysis

This section contains the following:

- Alignment of paired reads and filtering processes,

- Calculation of physical proximity values,
- Distance normalization and
- Statistical analyses for detecting significant associations.

1. Alignment of paired reads and filtering processes. The 36 bp paired reads were aligned by using Maq (http://maq.sourceforge.net/) with the setting of maximum outer distance (900 bp). Reference sequence of the fission yeast genome (20 090 706) was obtained from the Sanger Institute. Paired sequences containing HindIII sites at both ends of DNA molecules were maintained for subsequent analyses. In order to extract the data that reflect long-range associations, paired DNA sequences aligned to two genomic regions positioned <20 kb apart were discarded. To eliminate paired reads aligned to the repeat sequences, all the reads were aligned to the reference sequence using Bowtie (http://bowtie-bio.sourceforge.net/index.shtml) with the option –m1, which allows Bowtie to discard the sequences alignable to multiple positions. The discarded sequences were used to identify the paired reads derived from repetitive sequences. The paired sequences from repeats were removed from Maq-aligned data.

2. Calculation of physical proximity values. The entire fission yeast genome was divided into 20 kb sections. Paired sequences were assigned to two distant genomic sections according to positions of the reads. There are a total of 628 genomic sections. The total number of combinations between two sections was 196 878. All the paired reads were mapped to the genome. Total numbers of paired reads assigned to respective combinations were counted. Total counts of paired reads from the 3C sample were compared with those from RL control. Physical proximity value $I(i, j)$ was calculated as follows:

$$I(i,j) = N_{3C}(i,j)/N_{RL}(i,j).$$

$N_{3C}(i, j)$ indicates total count of paired reads from the 3C sample assigned to the combination between genomic loci i and j. $N_{RL}(i, j)$ is the count from RL control. $I(i, j)$ was discarded if $N_{RL}(i, j)$ was less than 4, because low values of $N_{RL}(i, j)$ appear to cause fluctuation in values of $I(i, j)$, resulting in 180 562 remaining combinations. The physical proximity values are accessible at the Wistar website (http://www.wistar.org/research_facilities/noma/pubdata.htm).

3. Distance normalization. Average physical proximity values between genomic loci separated by the same distances were gradually decreased along with the distances between two loci. Three curves for respective chromosomes were fitted by double-exponential curves. Physical proximity values were normalized by means of the following formula.

$$Q(i,j) = I(i,j)/F(x).$$

x represents the specific distance between two genomic loci i and j. $Q(i, j)$ is the distance-normalized value of physical proximity value, $I(i, j)$. $F(x)$ is the function indicating the fitting curves for respective chromosomes.

4. Statistical analyses for detecting significant associations. Statistical analyses were performed to test the hypothesis that genes related to some biological features associate together. For instance, the significance of associations among LTRs, highly and poorly expressed genes, cell-cycle regulated genes and genes in gene ontology groups was investigated. If associations among a specified group of genes are significant, the total physical proximity value among genomic sections containing those genes should be higher than that among randomly selected sections. According to this criterion, total physical proximity values among genomic sections containing genes in the target group were compared to those among genomic sections from a null model. For each target group, we calculated the total physical proximity value among the same number of genomic sections, randomly selecting from the entire genome. A null model was built by repeating this process 1000 times (1000 permutations). Distribution of total physical proximity values corresponding to the null model was used for the calculation of *P*-value.

(1) For the analyses of LTRs, highly and poorly expressed genes, and cell-cycle regulated genes, 80 sections from entire target genomic sections were randomly selected. The total physical proximity value was calculated as a sum of physical proximity values corresponding to all combinations among 80 sections. The distribution of total physical proximity values was built by 1000 permutations. The average of total physical proximity values was compared to the null model to test the significance of associations among target genomic sections.

(2) Only for M genes, 50 genomic sections, instead of 80 sections, were randomly selected, followed by statistical analyses as described in (1).

(3) For analyses of genomic sections containing particular gene arrangements, genes containing the same motifs, and genes in respective gene ontology groups, total physical proximity values corresponding to entire target genomic sections were calculated and compared to the null model.

Modeling the 3D genome structure

The modeling of the 3D genome structure was performed as described previously (14) with modifications. The fission yeast genome was modeled as strings of beads. Each bead displays a center of a 20 kb genomic section. There are a total of 622 beads covering the entire genome.

The first step was to calculate the 3D distance from the physical proximity value. Eighteen pairs of distant genomic loci were analyzed by FISH. Due to the distributions of FISH measurements, 30% of data points were truncated from both tails in order to remove possible outliers and the remaining middle 40% of the FISH data were used for the following calculations. The relationship between physical proximity values and FISH data was fitted by a non-linear regression curve. All physical proximity values were converted to 3D distances according to the fitted equation. The top 60% of physical

proximity values corresponding to 115 878 combinations were used for the following modeling processes.

The next step was to calculate coordinates of all beads separated by the distances calculated above. Let $p_i = (x_i, y_i, z_i)$ be the 3D coordinate of the *i*-th bead. $dist(p_i, p_j)$ denotes the Euclidean distance between p_i and p_j. Let $\delta_{i,j}$ be 3D distance converted from physical proximity values between two genomic sections *i* and *j*. All bead coordinates were finally found by minimizing the squared sum of differences between $dist(p_i, p_j)$ and $\delta_{i,j}$ as described by:

$$(p_1,...p_{622}) = \arg\min \sum_{i=1}^{622} \sum_{j=i+1}^{622} (dist(p_i,p_j) - \delta_{i,j})^2$$

This minimization was performed under the following five constraints:

(1) All beads must be present in a sphere with radius of $0.71\,\mu m$, which corresponds to a half of maximum distance observed by FISH. That is, for any $i = 1,...,622$,

$$dist(p_i, (0,0,0))^2 = x_i^2 + y_i^2 + z_i^2 \leq 0.71^2.$$

Without loss of generality, we set the origin (0,0,0) as the center of the sphere.

(2) The distance between adjoining beads must be within $0.133\,\mu m$ to $0.182\,\mu m$ (19).

$$0.133^2 \leq dist(p_i, p_{i+1})^2 \leq 0.182^2.$$

(3) To avoid overlap, all two beads must be positioned $>0.03\,\mu m$ apart (20).

$$dist(p_i, p_j)^2 \leq 0.03^2 \text{ for any } i \neq j.$$

(4) All three centromeres co-localize at the nuclear periphery. To reflect this, centromeres must lie within a small sphere with a radius of $0.03\,\mu m$ and the center of the sphere positioned at (0.68, 0, 0).

$$dist(p_c, (0.68,0,0))^2 = (x_c - 0.68)^2 + y_c^2 + z_c^2 \leq 0.03^2,$$

where p_c corresponds to the position of the centromere. There are three p_c representing three centromeres.

(5) Telomeres must localize at the nuclear periphery.

$$dist(p_t, (0,0,0))^2 = x_t^2 + y_t^2 + z_t^2 = 0.71^2,$$

where p_t corresponds to the position of the telomere. There are 6 p_t representing telomeres.

No constraints were applied for inter-chromosomal associations. Applying the above five constraints, the minimization was solved by AMPL software with IPOPT solver (21). The 3D structure of the entire fission yeast genome was built by smoothly interpolating the obtained 3D coordinates of the 622 beads. The modeled structure

was drawn by Pymol (22). The model structure is accessible at the Wistar website (http://www.wistar.org/research_facilities/noma/pubdata.htm).

FISH

FISH experiments were performed as described (23). To generate FISH probes, cosmid, plasmid or PCR-derived DNA fragments were labeled by incorporating Cy3-dCTP or Cy5-dCTP (GE Healthcare) using a random primer DNA labeling kit (Takara). Cosmid clones were obtained from the Sanger Institute. The cosmid cos212 and the plasmid pRS140 were used for preparing FISH probes specific to telomeres and centromeres, respectively. Stained cells were analyzed by a Zeiss Axioimager Z1 fluorescence microscope with oil immersion objective lens (Plan Apochromat, 100×, NA 1.4, Zeiss). Images were acquired at 0.2 μm intervals in the z-axis and deconvolved by Axiovision 4.6.3 software (Zeiss). More than 100 cells were analyzed for each experiment.

Expression analysis

Total RNA was extracted from cells as described previously (24). The total RNA sample (∼5 μg) was treated with 10 U of DNase I (Promega) at 37°C for 40 min, to remove contaminating genomic DNA and then purified by phenol/chloroform extraction. The resultant RNA sample was subjected to microarray analysis. Microarray experiments were conducted as described in the Nugen ovation manual and the Affymetrix genechip expression analysis technical manual. Briefly, 100 ng of total RNA was reverse transcribed by poly(T) nucleotides and cDNA was amplified by Ovation RNA amplification system v2 (Nugen Technologies). The amplified cDNA was biotinylated by Fl-ovation cDNA biotin module v2, followed by hybridization to Yeast genome 2.0 genechips (Affymetrix) at 45°C for 16 h. The array was washed with low (6× SSPE) and high (100 mM MES, 0.1 M NaCl) stringency buffers, and stained with streptavidin-phycoerythrin. Fluorescence signal was amplified by the addition of biotinylated anti-streptavidin and an additional aliquot of streptavidin–phycoerythrin stain. A confocal scanner was used to scan microarrays at excitation 570 nm. For initial data analysis, an Affymetrix command console was used to quantitate expression levels for targeted genes. Microarray data preprocessing, including normalization and background correction, was performed by the Mas5.0 software. The expression data have been deposited at NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) and are accessible through the GEO accession # GSE15108.

Based on the microarray data, expression levels were assigned to the genes. To analyze the relationship between expression levels of the genes and their associations, the entire fission yeast genome was first divided into 20 kb sections. There are 628 sections derived from the fission yeast genome. Each 20 kb section contains ∼10 genes. Expression levels of the 10 genes within a section were compared, and the maximum expression level from a gene was assigned to the section. The genomic sections were ranked by the expression levels.

The genomic sections corresponding to highly and poorly expressed genes were derived from the top 100 and bottom 100 sections, respectively.

Motif search

DNA sequences corresponding to 600 bp regions upstream of target genes were obtained from the Sanger Institute FTP server. Existing software such as MDSCAN, MEME, BioInspector, GLAM, Gibbs Motif Sampler, Weeder, Prority and SCOPE were first used to search for motifs with the setting allowing arbitrary length and wild cards. These methods did not produce any consistent motifs. Therefore, a new motif search method was developed. With our method, all possible combinations of 6–9 nt were searched exhaustively in the real data set and two background data sets (null models). The two null models were used to evaluate the significance of specific sequences identified in the real data set. One of the null models was DNA sequence derived from intergenic regions, while another null model was created by randomly shuffling sequence in the real data set so that the order of the nucleotides changes but the letter content was identical to the real set. Appearance frequency of the specific sequence in the real set was compared to those in the two null models. Sequence shuffling and obtaining intergenic sequence were repeated 1000 times. DNA sequences were recognized as motifs if appearance frequencies of the specific sequences were >21% of the total sequences in the real set and the ratios between appearance frequencies in the real set and the two null models were both greater than two. Finally, the motif with the highest score was selected and used to scan the TRANSFAC database to determine whether it had been previously identified (25).

RESULTS AND DISCUSSION

Capturing long-range associations throughout the fission yeast genome

In order to study the fission yeast genome organization, we have applied a modified Hi-C approach to our studies (Figure 1A). We first established that our 3C approach was suitable for analyzing the fission yeast genome organization by confirming the clustering of centromeres, as previously characterized by FISH analysis (Figure 1B) (26). The 3C sample contains hybrid DNA molecules reflecting physical associations between discrete genomic loci. In carrying out these studies, we developed a method, referred to as enrichment of ligation products (ELP), to prepare the 3C sequencing samples (Figure 1A). The ELP method involves an initial digestion of the 3C sample with a restriction enzyme (BfuCI) that recognizes a specific 4 bp sequence, followed by self-ligation and further treatment with another restriction enzyme used for 3C (HindIII in our experiment). As a result, the hybrid DNA fragments ligated together during the 3C experiment can be enriched for the sequencing step due to the linearity of these DNA molecules while non-linear spurious background sequences are reduced. The ELP-processed samples were then sequenced using an Illumina Genome Analyzer II with
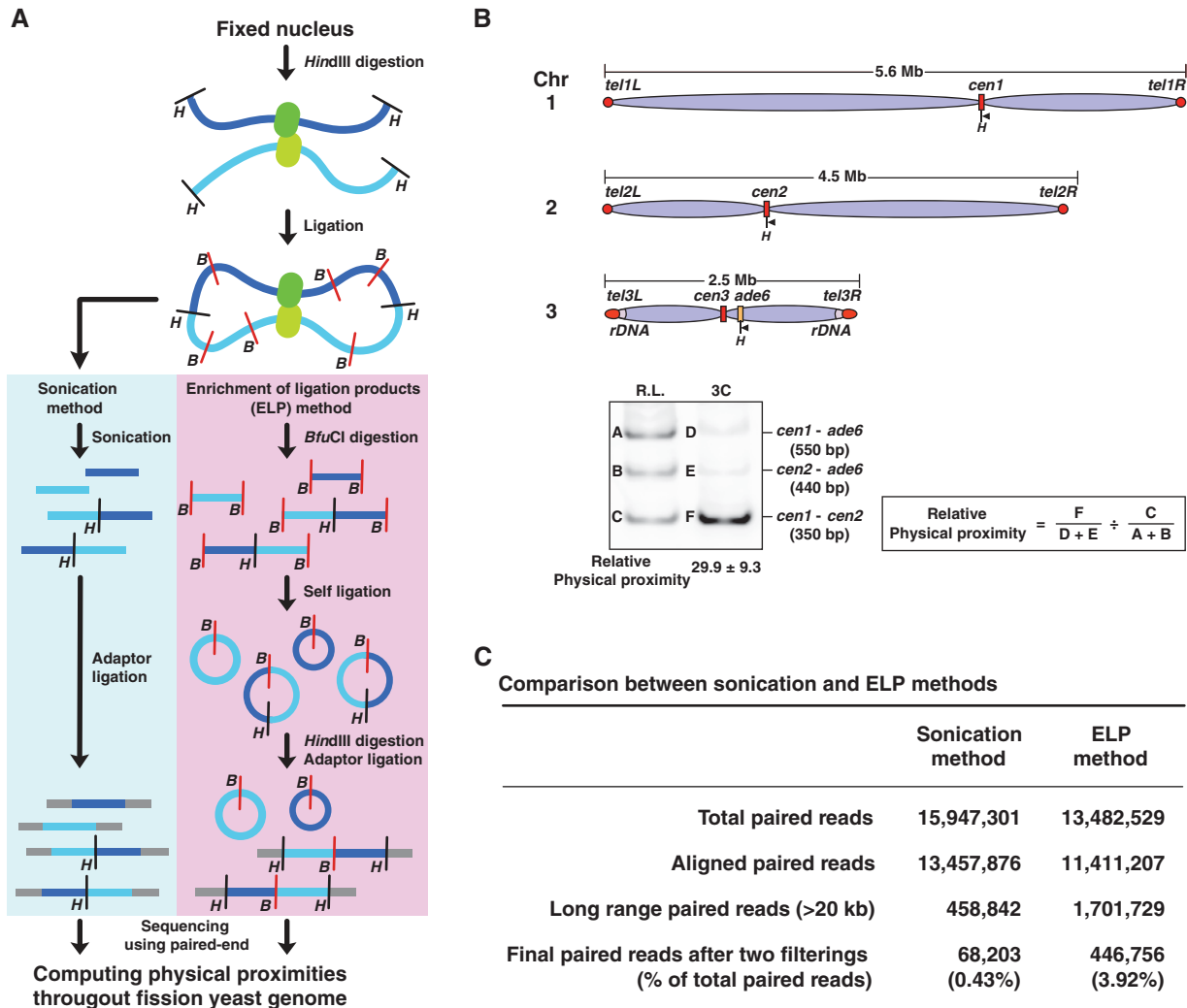
Figure 1. Capturing long-range associations between DNA fragments throughout the fission yeast genome. (**A**) Strategy of our genomics approach combining 3C and the massively parallel sequencing. The 3C procedure was followed by the ELP method, and the resultant sample was subjected to sequencing. (**B**) Detection of the association between centromeres 1 and 2 by 3C. The physical interaction between centromeres 1 and 2 was analyzed by 3C. The *ade6* gene locus serves as a negative control for association with centromeres. *H* indicates HindIII sites. The filled triangles show the locations and directions of the PCR primers used for 3C analysis. RL denotes the random ligation control sample. For the RL sample, un-cross-linked genomic DNA was cleaved by HindIII and randomly ligated by T4 DNA ligase. In contrast to the RL sample, the 3C sample reflects physical interaction between two distant genomic loci. Three primers were used for one PCR reaction. The bottom band in the 3C lane was relatively more intense than the two upper bands, suggesting an interaction between centromeres 1 and 2. 3C analyses were repeated three times, and the average value of relative physical proximity is shown beneath the representative result. This PCR-based assay shows reliability of the 3C procedure used for fission yeast, although global genome organization was from here on investigated based on the sequencing data. (**C**) Comparison between sequencing results derived from the ELP and sonication control samples. Approximately nine times more paired reads from the ELP sample compared to control remained after the filtering processes, indicating that the ELP method significantly enriches DNA fragments reflecting associations between distant genomic loci.

the Paired End (PE) module (Illumina, San Diego, CA, USA). The Paired End method determines the sequences present at both ends of the single DNA molecule, and can also determine whether or not they are derived from the same contiguous DNA fragment. Paired reads from >10 to 15 million DNA molecules were then mapped to genomic positions. Paired sequences found to be located <20 kb apart were first filtered out. The paired reads derived from repetitive DNA sequences such as centromeric repeats were also eliminated, even if only one end of the paired reads was assigned to repeats, because these sequences can be assigned to multiple genomic positions. The remaining sequences were then examined to identify

long-range genomic associations. Approximately a half million paired end reads derived from a single sequencing lane remained after the above filtering processes (Figure 1C). To obtain a sufficient number of paired reads to cover the entire fission yeast genome, the ELP-prepared sample was sequenced three times. In comparison to the simple application of the 3C sample to the massive sequencing, the ELP method resulted in an ~9-fold increase in the number of paired reads representing associations between genomic loci.

To check the reproducibility of our analysis, an independent 3C sample was processed using the ELP method and again sequenced three times. A randomly ligated (RL)

control sample, which does not reflect *in vivo* associations between genomic loci, was also processed with the ELP method and sequenced four times. Total numbers of paired reads after the filtering processes were 1.2–1.3 million for the 3C samples and 3.8 million for RL control (Supplementary Figure S1A). We found that paired reads from the RL control were not evenly distributed throughout the genome, indicating that there are obvious sequencing biases which also appeared to affect the distribution of paired reads from the 3C samples. We thus created our specific approach to determine a physical proximity value using normalization according to the distribution of paired reads from the RL control. We calculated a physical proximity value between distant 20 kb DNA fragments by comparing the total count of paired reads from the 3C sample with that from the RL control (Supplementary Figure S1B). The same calculation was carried out for every combination of 20 kb DNA fragments throughout the fission yeast genome. The physical proximity values from the two independent 3C samples (3C–1 and 3C–2) showed a clear correlation at a 20 kb resolution (Pearson's $r = 0.744$, $P < 2.2 \times 10^{-16}$), indicating that our methodology generates reproducible data (Supplementary Figure S1C). Resolutions at 10 and 40 kb indicated the lower and

higher correlations, respectively, compared to the 20 kb resolution (Supplementary Figure S1C). From here on we employ the 20 kb resolution data, judging from the correlation and size of genomic sections that was suitable for the following genomics analyses.

Verification of the physical proximity map by FISH

We plotted the physical proximity values throughout the three fission yeast chromosomes (Figure 2). The comprehensive map represents the physical proximity values between 20 kb DNA fragments distributed throughout the genome. We identified specific associations among centromeres and among telomeres. These genome structures are tightly linked to chromosome dynamics, and interactions were also detected by FISH analyses (Figure 3A and B) (26). In fission yeast, heterochromatin is distributed at centromeres, telomeres and a few other loci, and euchromatin is present in the remaining domains (17). It is known that RNAi machinery is involved in associations of these heterochromatic domains (27). We next tested whether other associations indicated in the map could also be detected using FISH analysis to visualize the intra-nuclear positioning of the various genomic loci. We investigated three combinations
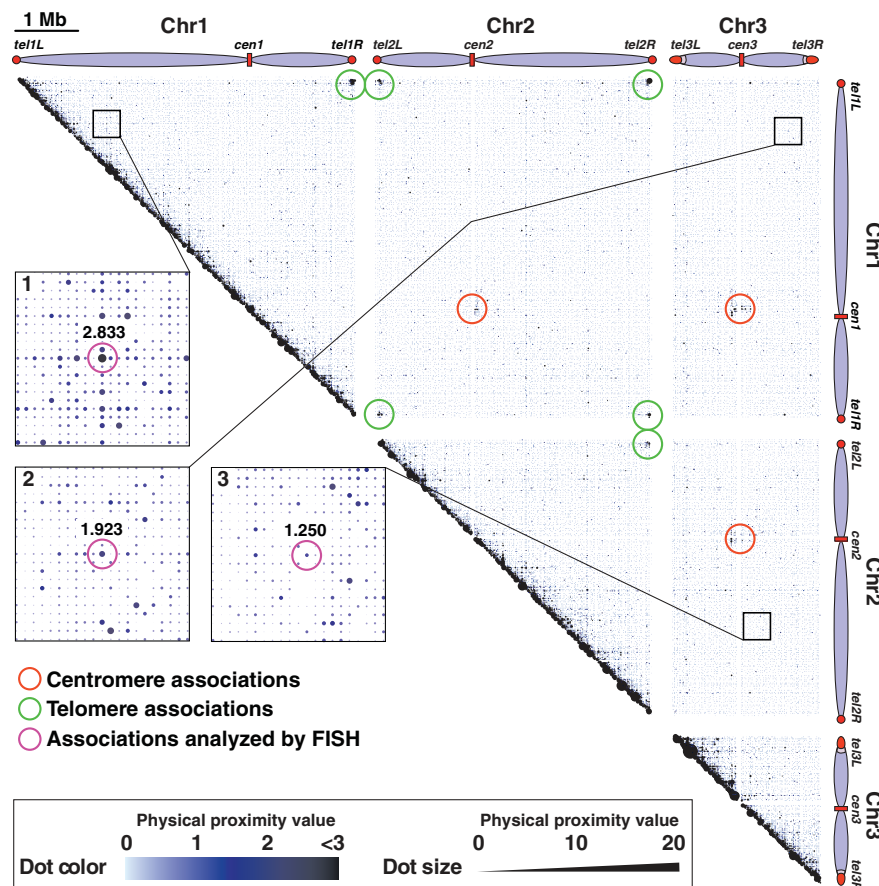

Figure 2. Comprehensive mapping of long-range associations throughout the fission yeast genome. Physical proximity values reflecting average association frequencies between 20 kb genomic sections in the cell population were calculated as described in Supplementary Figure S1 ('Materials and Methods' section) and plotted in the map. The physical proximity values noted in the enlarged views (1, 2 and 3) were verified by FISH (Figure 3C).
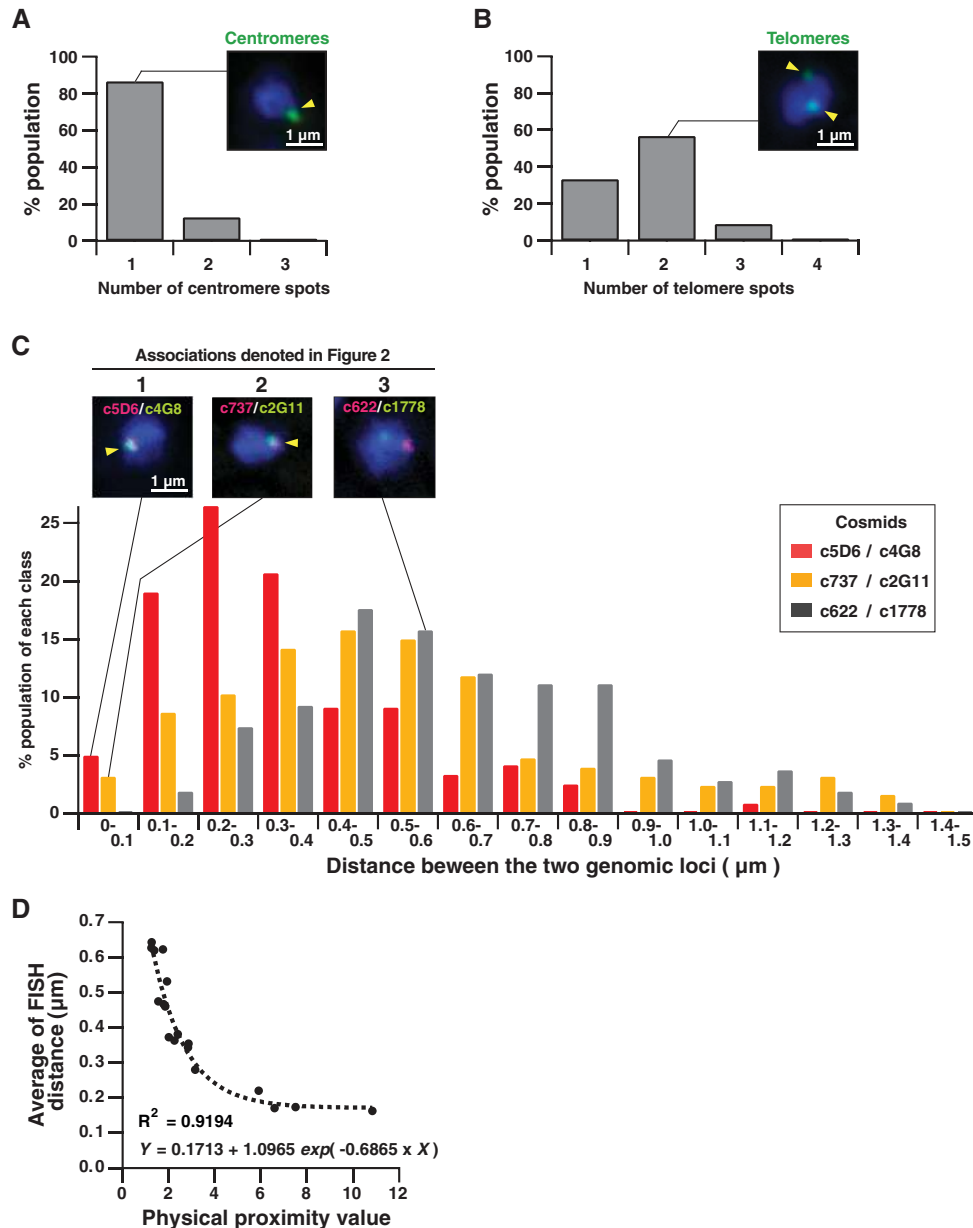
Figure 3. Verification of physical proximity values by FISH. (**A**) Visualizing centromeric clustering by FISH. FISH signal (green) visualizing centromeres 1, 2 and 3 often displayed a single spot, indicating the association among centromeres. Blue is DAPI signal. Typical image is shown on top. (**B**) Telomere association was visualized by FISH. FISH probe recognizes four telomeric regions in chromosomes 1 and 2. One or two FISH spots were frequently observed, indicating telomere associations. (**C**) The physical proximity values (1, 2 and 3) denoted in Figure 2 were evaluated by FISH. The two distant genomic loci (green and red) were visualized by FISH using the cosmid clones, and merged with DAPI signals (blue). Several images are shown on top. In each image, measurement of the distance between the genomic loci is indicated by linking the image with the histogram. The distance was measured between two focal centers. The percent populations of the observed distances between the genomic loci were binned into 0.1 μm. (**D**) Strong correlation between physical proximity values in the map and FISH data. Eighteen pairs of distant genomic loci were analyzed by FISH. The 18 combinations of genomic loci are annotated in Supplementary Table S1. Averages of distance values measured between two FISH spots were plotted along with physical proximity values between two genomic loci.

(1, 2 and 3) indicated in Figure 2, and found that the physical proximity values correlated with FISH data (Figure 3C). We performed extensive FISH analyses on a total of 18 combinations of genomic loci, and found the physical proximity values to be very strongly correlated with the FISH data ($R^2 = 0.9065$; Figure 3D). These observations support our interpretation that the physical proximity values in the map reflect global genome organization *in vivo*.

Modeling of the 3D genome structure

Three-dimensional structure of the budding yeast genome has recently been modeled using the Hi-C data (14). We employed a similar approach to model the fission yeast genome structure (See 'Materials and Methods' section). Physical proximity values were converted into 3D distances using the conversion formula obtained by comparing physical proximity values and FISH data for

18 pairs of genomic loci (Figure 3D and Supplementary Table S1). The 3D genome structure was modeled based on the calculated distances corresponding to 115 878 combinations between distant genomic loci (Figure 4A). Moreover, we validated the modeled genome structure by comparing the distances in the 3D structure to FISH data. Distances in the modeled structure and in FISH data lie near the 45° line ($R^2 = 0.8970$; Figure 4B), indicating that the modeled genome structure appears to reflect the *in vivo* structure to some extent. However, it is important to note that the modeled structure might not perfectly match the *in vivo* genome structure due to technical limitations. Physical proximity values used for the modeling of the genome structure only reflect average association
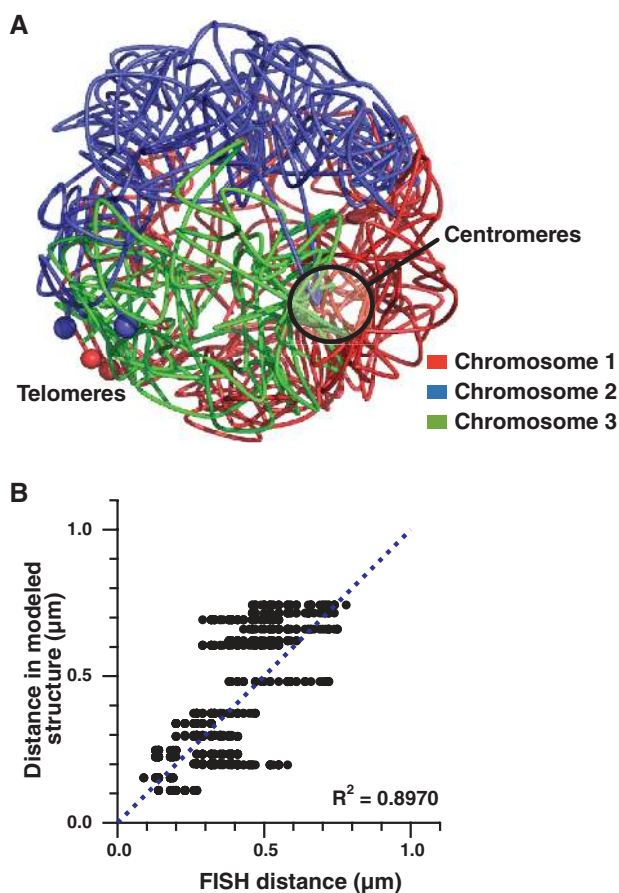

A

Centromeres

Telomeres

■ Chromosome 1
■ Chromosome 2
■ Chromosome 3

B

Figure 4. Modeled 3D structure of the fission yeast genome. (**A**) The 3D genome structure was modeled at a 20 kb resolution based on physical proximity values. Individual chromosomes are represented by different colors. Centromeres (open circle) in all three chromosomes are set to be clustered as an initial modeling constraint ('Materials and Methods' section). Telomeres in chromosomes 1 and 2 are highlighted by spheres. Telomeres in chromosome 3 consist of rDNA repeats and are known to be structurally distinct from telomeres in chromosomes 1 and 2. (**B**) Scatter plot of distances between two genomic loci in the modeled structure versus distances observed by FISH. Eighteen combinations of two distant genomic loci were analyzed by FISH (Figure 3D). FISH distances for each combination were measured in >100 cells. Due to the distributions of FISH measurements, 30% of data points were truncated from both tails in order to remove possible outliers and the remaining middle 40% of FISH data were plotted. The dotted lines represent the 45° line. The scatter plot indicated the high R^2-value (0.8970), supporting that the modeled genome structure strongly correlate with FISH data.

frequencies between genomic loci in the cell population, and do not directly represent stability of respective associations. For example, physical proximity values cannot distinguish between stable associations in a few cells and unstable associations in many cells. However, it is likely that stable associations such as telomere clustering occurs in many, if not all, cells, resulting in high scores of physical proximity values, which are major determinants for positioning of genomic loci in the modeled genome structure. This likely accounts for the modeled structure being strongly correlated with FISH data (Figure 4B).

In the modeled genome structure, we first noticed that the telomeres from chromosomes 1 and 2 were in close proximity, which was also indicated by FISH results (Figures 4A and 3B). This again suggests that the modeled genome structure at least partially reflects the *in vivo* structure. Interestingly, we also observed that three chromosomes were segregated into respective domains with overlapping junctions. This chromosome segregation partially results from the strong local associations that are represented diagonally in the physical proximity map (Figure 2). Those local associations between genomic loci separated by <1 Mb contribute to self-assembly of the respective chromosomes. Moreover, the average physical proximity value for intrachromosomal associations between genomic loci separated by >1.0 Mb was 0.64, while the average physical proximity value for interchromosomal associations was 0.59. This difference should not be observed when chromosomes are randomly disposed in the nucleus, supporting chromosome segregation in fission yeast. This disposition of chromosomes in the nucleus is similar to chromosome territories observed in mammalian cells (4,28). Our results are also consistent with previous observations, by which FISH analyses indicated chromosome territories existing in fission yeast (29). Together, our analyses suggest that the intra-nuclear disposition of the fission yeast chromosomes might to some extent be similar to the mammalian organization.

Physical proximity values negatively correlate with distances between genomic loci

We found strong local associations that are represented diagonally in the physical proximity map (Figure 2), most likely because those DNA fragments are relatively closely positioned in the nucleus. To examine the extent of the distance effect, we plotted the average ligation frequencies between genomic loci separated by the same distances, and found that the average ligation frequencies for the 3C sample were gradually decreased along with the distances, while the frequencies for the RL control samples were not related to the linear distances (Supplementary Figure S2). We also found that the average physical proximity values between genomic loci positioned less than ~1 Mb apart were gradually decreased along with the distances between two loci (Supplementary Figure S3A). The distance curves also revealed associations between left and right telomeres within the same chromosomes. Since it was possible that some local associations embedded in the map reflected specific local interactions, we tested this

possibility by using the distance curves to normalize the physical proximity values (Supplementary Figure S3B). This distance normalization eliminated a major population of local associations that likely resulted from random positioning of spatially linked genomic loci (Supplementary Figure S4A). Physical proximity values more than the average level ($\sim$1.0) imply that association frequencies between distant genomic sections are greater than the random association level. The distribution of physical proximity values indicated that 14 and 5% of the total combinations (180 562) between 20 kb genomic sections had values of more than 1.5 and 2.0, respectively (Supplementary Figure S4B). Associations that scored with physical proximity values >1.5–2.0 were likely to be detected by FISH in some populations of cells (Figure 3D).

Associations among LTR retrotransposons

We examined whether the distance-normalized map captures previously identified genome organizations. A previous study has shown that long-terminal repeat (LTR) retrotransposons cluster in the fission yeast nucleus (30). Our analysis also identified significant associations among DNA fragments containing LTRs ($P = 0.00529$, 1000 permutations; Figure 5A). Although paired reads derived from repetitive DNA sequences were removed by the filtering process as described above, we were able to investigate the associations among DNA fragments containing LTRs, because HindIII sites are not present within LTRs. We found that associations with physical proximity values >1.5 were increased by 4.2%, when associations between genomic sections containing LTRs were compared to the random control considering entire genomic sections. In other words, there were 493 (4.2%) additional associations derived from a total of 11 628 combinations between 153 LTR sections, as compared to the average association frequency between randomly picked genomic sections. This result again argues that the physical proximity map reflects the global genome organization *in vivo*. The physical proximity values are accessible at our website (See 'Materials and Methods' section) and can be used to identify novel genome organizations involving long-range associations. In the following sections, we exemplify how the physical proximity values can be used to investigate global genome organizations.

Associations among genomic sections containing specific gene arrangements

We examined whether gene arrangement influences association between genomic loci. We considered in total 36 gene arrangements involving 6 genes. Interestingly, genomic sections containing the specific gene arrangements tend to associate with one another in a statistically significant manner (Supplementary Figure S5). Genomic sections containing three consecutive convergent genes displayed the most significant association. Associations among genomic sections carrying two consecutive convergent genes were also significant. All the top 7 gene arrangements contained consecutive convergent genes, but

the remaining 29 gene arrangements did not have any consecutive convergent genes. The gene arrangement without any convergent genes displayed the lowest average physical proximity value. These results suggest that consecutive arrangement of convergent genes is favored for associations between genomic regions. It has been shown that cohesin is recruited to convergent genes in fission yeast (31,32). In mammals, cohesin is implicated in association between genomic loci (33–37). It is possible that, in fission yeast, cohesin might be involved in the association between genomic regions containing consecutive convergent genes. In any case, our analyses suggest that gene arrangements contribute to global genome organization in fission yeast.

Associations among highly expressed genes

To explore the influence of transcription on global genome organization, we asked whether genomic sections containing highly expressed genes associate in the nucleus. Our analysis revealed significant associations between genomic regions containing highly expressed genes, as compared to randomly selected genes serving as a control ($P = 0.0252$, 1000 permutations; Figure 5B). Associations that scored with physical proximity values of more than 1.5 were increased by 3.5% (172 combinations) when associations between highly expressed genes were compared to the random control. In clear contrast, associations among the poorly-expressed genes were not different from the control ($P = 0.418$, 1000 permutations; Figure 5B), suggesting that highly transcribed genes tend to associate with one another in a statistically significant manner. It has been shown that active genes are co-localized to the shared nuclear sites referred to as transcription factories in mammalian cells, although the exact functions of transcription factories and their assembly processes are still unclear (6,7,38–40). Our results suggest that highly active genes frequently co-localize at transcription factories or functionally similar entities present in the fission yeast nucleus.

Associations among co-regulated genes during the cell-cycle progression

We next examined whether co-regulated genes associate in the fission yeast nucleus. It has been reported that many genes in fission yeast are periodically regulated during the cell cycle (41). Those periodically transcribed genes were previously classified into four groups representing expression peaks during M, G1, S or G2 phases. Interestingly, we found that only G2 phase genes exhibited significant associations ($P = 0.0285$, 1000 permutations), whereas genes in the other groups did not show significant associations (Figure 5C). Association frequencies among G2 genes were similar to those among LTR retrotransposons (Figure 5A and C). Associations scored with physical proximity values of more than 1.5 were increased by 3.8% (259 combinations) when associations among G2 genes were compared to the random control. It is noteworthy that the 3C samples were prepared from asynchronous cultures, which predominantly contain G2 cells ($\sim$75%). Therefore, our data suggest that in fission yeast,
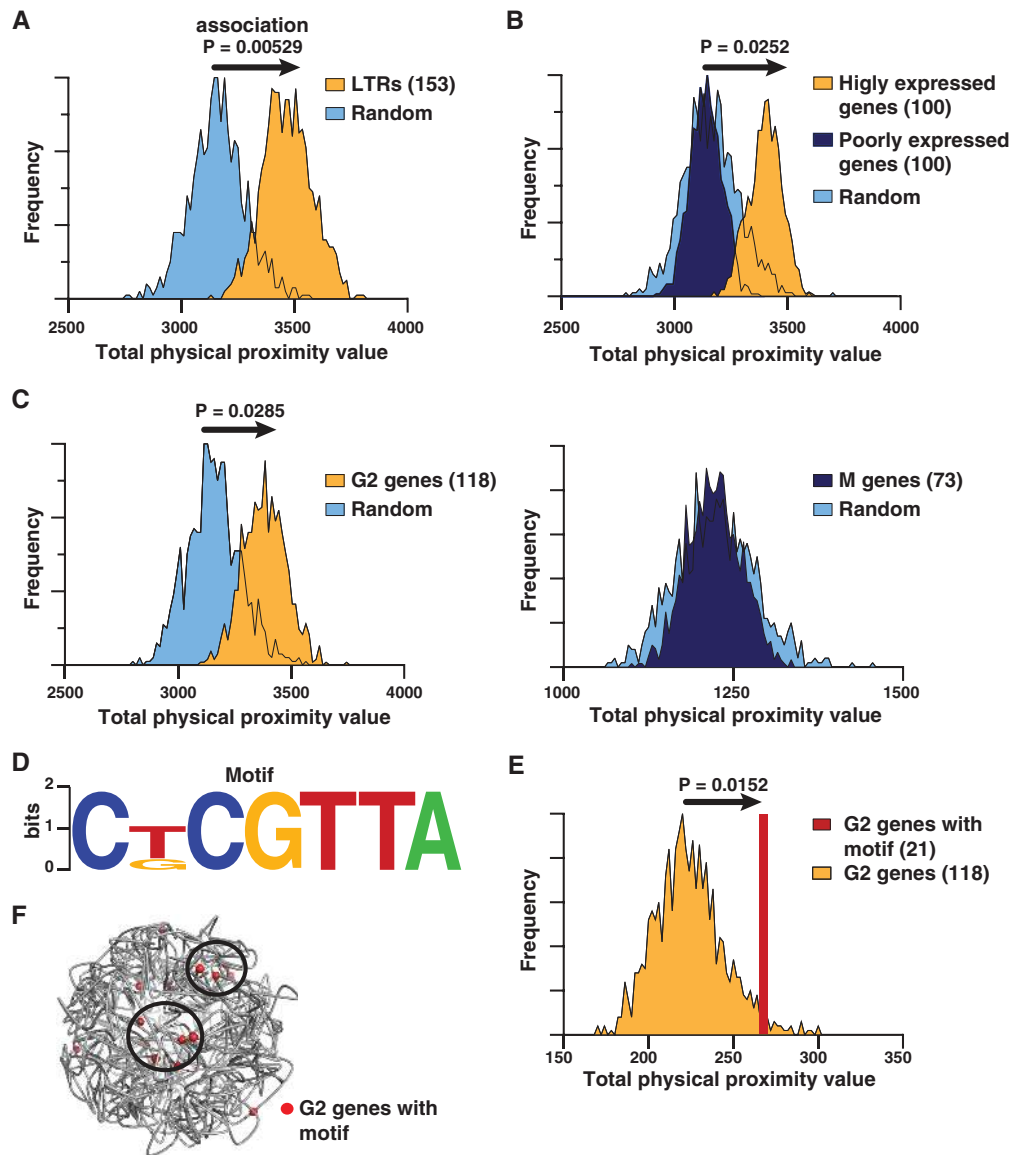
Figure 5. Significant associations among highly expressed genes and co-regulated genes during the cell cycle. (**A**) Associations among genomic sections containing LTR derived from retrotransposons. Figures in parenthesis indicate the number of 20 kb genomic sections containing LTR. The total physical proximity value was calculated as a sum of physical proximity values corresponding to all combinations among 80 genomic sections randomly selected from 153 sections containing LTR. Distribution of total physical proximity values was built by 1000 permutations. For the random control, 80 genomic sections were randomly picked from entire genomic sections, followed by 1000 permutations. The statistical analysis to test the significance of associations among LTRs (yellow) compared to random control (light blue) is described in 'Materials and Methods' section. (**B**) Highly expressed genes, but not poorly expressed genes, tend to associate together. The highly expressed and poorly expressed gene sections were classified as described in 'Materials and Methods' section. The statistical analyses were carried out as described in A. (**C**) G2 genes showing the expression peak during G2 phase tend to associate with one another (left), whereas association frequencies among M genes did not differ from the random control (right). The statistical analyses for G2 genes were carried out as described in (A). For M genes, 50 genomic sections were randomly selected from 73 genomic sections containing M genes, followed by 1000 permutations. (**D**) A new DNA motif identified at the upstream regions of G2 genes. Twenty-one of a total of 118 G2 genes contained the motif. This motif did not match the previously identified motifs listed in the TRANSFAC database (25). (**E**) G2 genes containing the motif showed significantly enhanced associations compared to the associations among entire G2 gene members. The red bar indicates the total physical proximity value among 21 G2 genes containing the motif. To test the significance of associations among G2 genes containing the motif, 21 G2 genes were randomly selected from a total of 118 G2 genes, and total physical proximity values were calculated. Distribution of 1000 permutations was plotted (yellow) and used for the calculation of the *P*-value. Associations that scored with physical proximity values more than 1.5 are listed in Supplementary Table S2. (**F**) Motif-containing G2 genes (red spheres) are shown in the modeled genome structure. The several motif-containing G2 genes were positioned in proximity (open circles).

G2 genes tend to associate with one another when activated. Since a majority of the cells in the culture are in G2, it is possible that other underrepresented cell-cycle-regulated genes associated with M, G1 and S phase might also associate during their respective cell-cycle stages, although this requires further experimental validation.

The regulation of periodically expressed genes involves interaction with specific transcription factors (41). In examining the upstream sequence of the G2 genes, we

have identified a new sequence motif, C[T/G]CGTTA, within the 600 bp region upstream of 21 G2 genes (Figure 5D). The motif was frequently positioned between the transcription start site and 200 bp upstream. Remarkably, G2 genes with this motif showed significantly stronger associations compared to associations among the entire G2 gene group ($P = 0.0152$, 1000 permutations; Figure 5E), suggesting that an unidentified DNA binding protein, likely a transcription factor recognizing this G2 gene-related motif, may facilitate these associations. Consistent with this result, we found that several G2 genes containing the motif were present in proximity in the modeled genome structure (Figure 5F). Moreover, almost all G2 genes (107/118 G2 genes) contain a degenerate motif with one mutation in the perfect motif. It is possible that the degenerate motifs in G2 genes might be less tightly bound by the potential factors, causing significantly enhanced associations among the entire G2 gene population compared to the random control (Figure 5C). It has recently been reported that in mouse, co-regulated genes preferentially cluster at transcription factories, and that this clustering is mediated by binding of the transcription factor Klf1 to the genes (8). Therefore, our data suggest that co-regulated genes in fission yeast associate with one another in a fashion functionally similar to the mammalian transcription factories.

Associations among genes in gene ontology groups

Our analyses had suggested that co-regulated genes significantly associate with one another. We next expanded our study to the entire gene population and asked whether genes involved in other particular biological process also frequently associate. To this end, we investigated the significance of the associations among a group of annotated genes classified by gene ontology in the fission yeast genome database (42,43). We analyzed 467 gene ontology groups containing 26–121 genes in 20–100 genomic sections. This range of genomic sections was chosen to avoid a high false-negative rate. We observed that genes from 23 gene ontology groups showed the significant associations compared to the random controls (Figure 6A). We discarded 6 out of the 23 gene ontology groups, because they were obviously subgroups of other main groups, which also showed significant associations. The remaining 17 gene ontology groups included metabolic process, transmembrane transporter activity, response to stimulus, regulation of Ras-GTPase activity and cell wall biogenesis.

If genes in the respective gene ontology groups associate through binding of transcription factors, then comparative sequence analyses should find conserved DNA motifs at the promoter regions. Indeed, we found new conserved DNA motifs present at the promoter regions of the genes in the four ontology groups (Figure 6B). More importantly, those genes containing these DNA motifs showed significantly enhanced associations compared to associations among the entire gene members in the respective gene ontology groups (Figure 6B). In agreement with these results, we found that several motif-containing genes in the cellular carbohydrate catabolic process were

present in proximity in the modeled genome structure (Figure 6C). We also found a similar positioning of motif-containing genes derived from the three other gene ontology groups in the modeled structure. These results suggest the importance of the DNA motifs and the potential involvement of factors binding to those motifs in facilitating associations between functionally defined genes in particular gene ontology groups.

We next investigated whether genes in many gene ontology groups might weakly associate with one another. To test this possibility, we plotted the distribution of average physical proximity values for 465 gene ontology groups and compared it to the distribution of the values for hypothetical random groups (Figure 6D). Distribution of average physical proximity values for actual gene ontology groups was significantly shifted to the right (Kolmogorov-Smirnov test $P = 1.96 \times 10^{-54}$). Average physical proximity values of most of the gene ontology groups (97%) represented more than 0.9, whereas only about half (58%) of the hypothetical groups had more than 0.9, suggesting that genes in many gene ontology groups tend to weakly associate with one another. It has recently been suggested that individual genes are confined to the distinct subnuclear compartments, referred to as gene territories in budding yeast (44). It is possible that genes in many gene ontology groups might be present at shared gene territories, although future study is essential to infer any biological functions related to the weak associations observed among genes in many gene ontology groups.

DNA motif-dependent genome organization in fission yeast

We have demonstrated that highly transcribed genes, co-regulated genes, and genes from particular gene ontology groups tend to co-localize in the *in vivo* genome structure. The associations among highly transcribed genes are reminiscent of the transcription factories proposed to exist in mammals, although the functional role of such transcription factories remains unclear (4–7). It has been recently suggested that the transcription factor Klf1 is involved in the association of genes with transcription factories in mouse (8). Our study indicated that genes containing the same DNA motifs at promoter regions associate with one another in the significantly enhanced frequencies, suggesting that unknown factors, likely transcription factors, play a role in gene associations. The DNA motif-dependent gene associations were observed for co-regulated genes during the cell cycle as well as functionally defined genes in particular gene ontology groups. Our current hypothesis is that transcription factors binding to the motifs are involved in the functional organization of the global genome structure, which is suitable for coordinated expression of genes dispersed throughout the genome. Future studies that attempt to address the mechanism of DNA motif/transcription factor-mediated gene associations should lead to new insights into complex genome wide processes in functional genome organization coupled with transcriptional regulation.
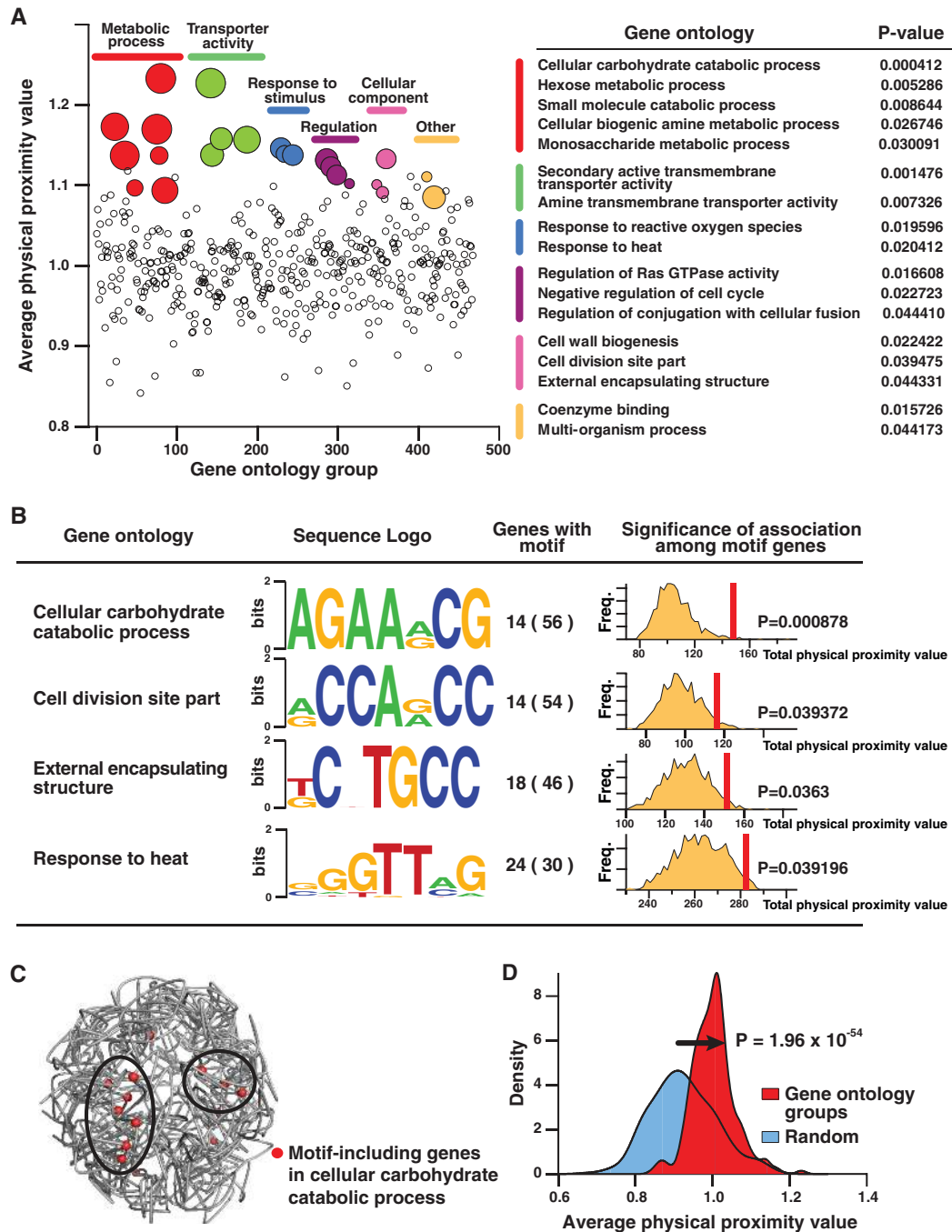
Figure 6. Significant associations among genes in gene ontology groups. (**A**) Genes in 23 gene ontology groups displayed significant associations. The average physical proximity values for 467 gene ontology groups were plotted in the graph (left). Average physical proximity value was calculated as total physical proximity value divided by total number of combinations among every genomics section containing genes in each gene ontology group. Significance of associations among genes in respective gene ontology groups was tested as described in 'Materials and Methods' section. Twenty-three gene ontology groups revealing significant associations are shown by filled circles. Several gene ontology groups derived from the same parent gene ontology groups are represented in the same colors. The size of the circles correlates with the significance of gene associations. Gene ontology groups revealing significant gene associations are listed along with the *P*-values (right). Six gene ontology groups were removed from 23 initially identified gene ontology groups, because they shared a major population of genes with other groups. Descriptions of respective gene ontology groups showing significant associations are provided in Supplementary Table S3. (**B**) Enhanced associations among genes in the particular gene ontology groups through new DNA motifs. All 17 gene ontology groups showing significant associations were subjected to the DNA motif analyses and four significant motifs were identified ('Materials and Methods' section). These motifs did not match the previously identified motifs listed in the TRANSFAC database (25). Numbers of genes contained the motifs are shown next to the motifs. Figures in parentheses indicate total numbers of genes in respective gene ontology groups. The statistical test was carried out as described in Figure 5E. The total physical proximity values among motif-including genes (red) were compared to distributions of values derived from entire gene members in respective gene ontology groups (filled yellow). Associations that scored with physical proximity values >1.5 are listed in Supplementary Table S2. (**C**) Positions of motif-including genes in cellular carbohydrate catabolic process (red spheres) are shown in the modeled genome structure. The several motif-including genes were present in the vicinity (open circles). (**D**) Increase in population showing weak associations among genes in gene ontology groups compared to the random control. Distribution of average physical proximity values of the actual 467 gene ontology groups (red) was compared to distribution derived from 467 hypothetical groups (blue) containing the same number of randomly selected genes.

REFERENCES

1. Misteli,T. (2007) Beyond the sequence: cellular organization of genome function. *Cell*, **128**, 787–800.
2. Iwasaki,O., Tanaka,A., Tanizawa,H., Grewal,S.I. and Noma,K. (2010) Centromeric localization of dispersed Pol III genes in fission yeast. *Mol. Biol. Cell.*, **21**, 254–265.
3. Thompson,M., Haeusler,R.A., Good,P.D. and Engelke,D.R. (2003) Nucleolar clustering of dispersed tRNA genes. *Science*, **302**, 1399–1401.
4. Fraser,P. and Bickmore,W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature*, **447**, 413–417.
5. Jackson,D.A., Hassan,A.B., Errington,R.J. and Cook,P.R. (1993) Visualization of focal sites of transcription within human nuclei. *EMBO J.*, **12**, 1059–1065.
6. Sutherland,H. and Bickmore,W.A. (2009) Transcription factories: gene expression in unions? *Nat. Rev. Genet.*, **10**, 457–466.
7. Dorier,J. and Stasiak,A. (2010) The role of transcription factories-mediated interchromosomal contacts in the organization of nuclear architecture. *Nucleic Acids Res.*, doi, 10.1093/nar/gkq666.
8. Schoenfelder,S., Sexton,T., Chakalova,L., Cope,N.F., Horton,A., Andrews,S., Kurukuti,S., Mitchell,J.A., Umlauf,D., Dimitrova,D.S. *et al.* (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.*, **42**, 53–61.
9. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.
10. Dostie,J., Richmond,T.A., Arnaout,R.A., Selzer,R.R., Lee,W.L., Honan,T.A., Rubio,E.D., Krumm,A., Lamb,J., Nusbaum,C. *et al.* (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.*, **16**, 1299–1309.
11. Simonis,M., Klous,P., Splinter,E., Moshkin,Y., Willemsen,R., de Wit,E., van Steensel,B. and de Laat,W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.*, **38**, 1348–1354.
12. Zhao,Z., Tavoosidana,G., Sjolinder,M., Gondor,A., Mariano,P., Wang,S., Kanduri,C., Lezcano,M., Sandhu,K.S., Singh,U. *et al.* (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.*, **38**, 1341–1347.
13. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
14. Duan,Z., Andronescu,M., Schutz,K., McIlwain,S., Kim,Y.J., Lee,C., Shendure,J., Fields,S., Blau,C.A. and Noble,W.S. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**, 363–367.
15. Rodley,C.D., Bertels,F., Jones,B. and O'Sullivan,J.M. (2009) Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genet. Biol.*, **46**, 879–886.
16. Wood,V., Gwilliam,R., Rajandream,M.A., Lyne,M., Lyne,R., Stewart,A., Sgouros,J., Peat,N., Hayles,J., Baker,S. *et al.* (2002) The genome sequence of schizosaccharomyces pombe. *Nature*, **415**, 871–880.
17. Cam,H.P., Sugiyama,T., Chen,E.S., Chen,X., FitzGerald,P.C. and Grewal,S.I. (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.*, **37**, 809–819.
18. Noma,K., Cam,H.P., Maraia,R.J. and Grewal,S.I. (2006) A role for TFIIIC transcription factor complex in genome organization. *Cell*, **125**, 859–872.
19. Bystricky,K., Heun,P., Gehlen,L., Langowski,J. and Gasser,S.M. (2004) Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl Acad. Sci. USA*, **101**, 16495–16500.
20. Dehghani,H., Dellaire,G. and Bazett-Jones,D.P. (2005) Organization of chromatin in the interphase mammalian cell. *Micron*, **36**, 95–108.
21. Wächter,A. and Biegler,L.T. (2005) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.*, **106**, 25–57.
22. Schrodinger,L.L.C. (2010) *The PyMOL Molecular Graphics System*, New York, Version 1.2r1.
23. Sadaie,M., Naito,T. and Ishikawa,F. (2003) Stable inheritance of telomere chromatin structure and function in the absence of telomeric repeats. *Genes Dev.*, **17**, 2271–2282.
24. Volpe,T.A., Kidner,C., Hall,I.M., Teng,G., Grewal,S.I. and Martienssen,R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.
25. Matys,V., Fricke,E., Geffers,R., Gossling,E., Haubrock,M., Hehl,R., Hornischer,K., Karas,D., Kel,A.E., Kel-Margoulis,O.V. *et al.* (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, **31**, 374–378.
26. Funabiki,H., Hagan,I., Uzawa,S. and Yanagida,M. (1993) Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J. Cell Biol.*, **121**, 961–976.
27. Hall,I.M., Noma,K. and Grewal,S.I. (2003) RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl Acad. Sci. USA*, **100**, 193–198.
28. Lanctot,C., Cheutin,T., Cremer,M., Cavalli,G. and Cremer,T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.*, **8**, 104–115.
29. Scherthan,H., Bahler,J. and Kohli,J. (1994) Dynamics of chromosome organization and pairing during meiotic prophase in fission yeast. *J. Cell Biol.*, **127**, 273–285.
30. Cam,H.P., Noma,K., Ebina,H., Levin,H.L. and Grewal,S.I. (2008) Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature*, **451**, 431–436.
31. Gullerova,M. and Proudfoot,N.J. (2008) Cohesin complex promotes transcriptional termination between convergent genes in S. *pombe*. *Cell*, **132**, 983–995.
32. Schmidt,C.K., Brookes,N. and Uhlmann,F. (2009) Conserved features of cohesin binding along fission yeast chromosomes. *Genome Biol.*, **10**, R52.
33. Hadjur,S., Williams,L.M., Ryan,N.K., Cobb,B.S., Sexton,T., Fraser,P., Fisher,A.G. and Merkenschlager,M. (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, **460**, 410–413.

34. Hou,C., Dale,R. and Dean,A. (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl Acad. Sci. USA*, **107**, 3651–3656.

35. Mishiro,T., Ishihara,K., Hino,S., Tsutsumi,S., Aburatani,H., Shirahige,K., Kinoshita,Y. and Nakao,M. (2009) Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.*, **28**, 1234–1245.

36. Nativio,R., Wendt,K.S., Ito,Y., Huddleston,J.E., Uribe-Lewis,S., Woodfine,K., Krueger,C., Reik,W., Peters,J.M. and Murrell,A. (2009) Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.*, **5**, e1000739.

37. Wood,A.J., Severson,A.F. and Meyer,B.J. (2010) Condensin and cohesin complexity: the expanding repertoire of functions. *Nat. Rev. Genet.*, **11**, 391–404.

38. Chakalova,L., Debrand,E., Mitchell,J.A., Osborne,C.S. and Fraser,P. (2005) Replication and transcription: shaping the landscape of the genome. *Nat. Rev. Genet.*, **6**, 669–677.

39. Cook,P.R. (2010) A model for all genomes: the role of transcription factories. *J. Mol. Biol.*, **395**, 1–10.

40. Osborne,C.S., Chakalova,L., Brown,K.E., Carter,D., Horton,A., Debrand,E., Goyenechea,B., Mitchell,J.A., Lopes,S., Reik,W. *et al.* (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.*, **36**, 1065–1071.

41. Rustici,G., Mata,J., Kivinen,K., Lio,P., Penkett,C.J., Burns,G., Hayles,J., Brazma,A., Nurse,P. and Bahler,J. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.*, **36**, 809–817.

42. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.

43. Aslett,M. and Wood,V. (2006) Gene ontology annotation status of the fission yeast genome: preliminary coverage approaches 100%. *Yeast*, **23**, 913–919.

44. Berger,A.B., Cabal,G.G., Fabre,E., Duong,T., Buc,H., Nehrbass,U., Olivo-Marin,J.C., Gadal,O. and Zimmer,C. (2008) High-resolution statistical mapping reveals gene territories in live yeast. *Nat. Methods*, **5**, 1031–1037.